**4. Data, Methods, and Summary Statistics**

*4.1 Data*

I gather institutional and insider ownership data on U.S. listed companies from CDA / Spectrum Compact Disclosure for each year from 1990 to 2005. I exclude financial firms and utilities because they are highly regulated by the government. The ownership data is then merged with Compustat data. The final sample consists of 10,668 firms and

79,890 firm-years. Some firms are missing data or not present in the sample for enough firm-years to perform certain analysis. In such cases, these firms are not used.

Annual dividends and stock repurchases are measured in dollars and scaled by the dollar book value of assets. Repurchases are defined as the dollar amount of stock repurchases minus the dollar amount of stock issues. If stock issues are greater than stock repurchases, the repurchase amount is set to zero. Changes in repurchases are measured as the repurchases of the current year minus repurchases of the previous year, divided by the book value of assets from the previous year. Changes in dividends are measured similarly. Total payout is defined as the sum of the dollar value of common dividends and repurchases.

Fama and French (2001) find in a study of U.S. firms that dividends are trending through time. They also find that firm profitability, size and growth opportunities are related to dividends. Therefore, I control for differences across firms using variables that control for these relationships. I use earnings before interest and taxes scaled by total assets as a proxy for profitability. Size is controlled for by using the logarithm of market value and the logarithm of revenue. I use *q* to control for growth opportunities. Following Dlugosz, Fahlenbrach, Gompers, and Metrick (2006), I calculate the variable *q* as the ratio of the market value of assets to the book value of assets where market value is calculated as the sum of the book value of assets and the market value of common stock less the book value of common stock and deferred taxes. All regressions include dummy variables for each year of the data sample to control for time effects on the relationship between institutional ownership and payouts.

DeAngelo, DeAngelo, and Stulz (2006) report a strong association between a company’s earned/contributed equity mix, which they use as a proxy for the life-cycle stage of a company, and dividends. Therefore, I follow them by using the earned/contributed equity mix defined as retained earnings to the book value of total equity to control for the life-cycle stage of the firm. Firm stock turnover is included as a control because Banerjee, Gatchev, and Spindt (2007) find that turnover is related to dividends. Jensen (1986) proposes debt can substitute for dividends, so firm debt to asset ratio is included.

I use net income plus depreciation and amortization minus capital expenditures as a measure of cash flow. Notably, this cash flow measure does not subtract dividends or repurchases as many measures of cash flow do. This is done to simplify the analysis of dividends, repurchases or payouts as a percentage of free cash flow. I divide this cash flow measure by total book value of assets to provide scale. The detailed definitions of all variables are shown in Table 1.

Table 1. Variable Definitions

Variable Description Definition

Panel A: Summary Statistics and Correlation Table Variables

|  |  |  |
| --- | --- | --- |
| *N* | Number of Firms | The number of firms. |
| *Inst* | Institutional Ownership | The fraction of shares owned by institutions. |
| *MktCap* | Market Capitalization | The dollar market value of common stock in millions. |
| *LifeCycle* | Firm Life-cycle | The ratio of retained earnings to total equity. |
| *q* | Investment Opportunities | Market value of assets to the book value of assets |
| *CashFlow* | Free Cash Flow | Free cash flow to total assets. |
| *Div* | Dividend Ratio | Dividends to book value of assets. |
| *Payout* | Payout Ratio | Total payout divided by book value of assets. |
| *Repurch* | Stock Repurchase Ratio | Stock repurchases to book value of assets. |
| *PayIncr* | Payout Increases | The percentage of firms which increased their total  payout per share. |
| *PayDecr* | Payout Decreases | The percentage of firms which decreased their total payout per share. |
| *RepIncr* | Stock Repurchase Increases | The percentage of firms which increased their repurchases per share. |
| *RepDecr* | Stock Repurchase  Decreases | The percentage of firms which increased their repurchases per share. |

Panel B: Regression Dependent Variable(Measured as changes in values from year *t* – 1 to *t*.)

*Payout* Payout Ratio Total payout divided by book value of assets.

Panel C: Regression Independent Variables (Measured as changes in values from year *t* – 2 to *t* - 1.)

|  |  |  |
| --- | --- | --- |
| *Inst* | Institutional Ownership | The fraction of shares owned by institutions. |
| *CashFlow* | Free Cash Flow | Free cash flow to total assets. |
| *q* | Investment Opportunities | Market value of assets to the book value of assets |
| *Debt* | Debt Ratio | Debt to assets. |
| *Turnover* | Stock Turnover | Firm common stock turnover. |
| *LifeCycle* | Firm Life-cycle | The ratio of retained earnings to total equity. |
| *MktCap* | Market Capitalization | The dollar market value of common stock in millions. |
| *ROA* | Return on Assets | Earnings before interest and taxes divided by total assets. |
| *Insider* | Insider Ownership | The fraction of shares owned by insiders. |
| *Insider2* | Insider Ownership Squared | The squared value of Insider. |
| *Revenue* | Revenue | The logarithm of firm revenue. |

All data is yearly data from 1990-2005. Institutional and insider ownership data from CDA / Spectrum

Compact Disclosure. All other data is from Compustat. Detailed descriptions in article text.

*4.2 Methods*

If there is a relationship between institutional investors and payouts, it is difficult to discern if institutional investors influence payouts or if payouts influence institutional investors or both. Therefore, I adopt a regression methodology which accounts for endogeneity and establishes causality.

To help address this causality issue, I run regressions on changes in dependent variables from year *t* – 1 to *t* on changes in independent variables from *t* – 2 to *t* – 1 to establish causality. All regressions use firm fixed effects. Firm fixed

effect regressions are useful because they control for all stable characteristics of a firm (including industry), whether measured or not. This appealing feature of firm fixed effects regressions combined with the use of yearly dummy variables to control for time-varying omitted characteristics helps to control for endogeneity issues in my analysis. Using the yearly dummy variables with fixed effects effectively gives each year its own intercept. Intercepts in fixed effects regressions are calculated as an average value of the unobserved fixed effects for each firm. The intercept values and yearly intercept values are not relevant to my analysis. Therefore, they are not reported in my regression results.

For robustness and to further address potential endogeneity, I use a difference generalized method of moments (GMM) methodology that is based on the methodology employed in Holtz-Eakin, Newey, and Rosen (1988) with refinements and validity tests developed by Arellano and Bond (1991). I implement the methodology using the Stata command xtabond2. I use methods described in Roodman (2009).

Difference GMM removes fixed effects and uses lagged values of the dependent variable and independent variables of interest as instruments. This method avoids endogeneity problems associated with using fixed-effects when there is autocorrelation in the dependent variable. It also corrects for any concurrent endogeneity problems associated with the inclusion of lagged independent variables.

Many notable research papers have used difference GMM in their analysis, including Brown, Fazzari, and Petersen (2009) and Brossard et al. (2013). Almeida, Campello, and Galvao (2010) assess the performance of difference GMM and find that its results conform to theoretical expectations in regressions that use data which contains firm fixed effects and heteroskedasticity.

This methodology is ideal for use in panel samples with a limited number of time periods and a large number of firms. My data consists of a maximum of 16 years of data for over 10,000 firms. Difference GMM is also designed to be implemented in situations with the following characteristics: a dependent variable that depends on past realizations of itself, independent variables that are not strictly exogenous, and firm fixed effects (Roodman, 2009). If conceptually and statistically sound instruments for endogenous independent variables are available, firm fixed effects regressions using those instruments would be preferable to using difference GMM. Unfortunately, I was unable to find valid instruments. Difference GMM uses lags of the endogenous regressors as instruments. This shrinks the size of the dataset because at least one year of data has to be dropped for each firm. In my implementation of difference GMM, only one year has to be dropped for each firm.

The dependent variables in my regressions depend on past realizations of the dependent variable because current payout policy is largely dependent on past payout policy. In my robustness checks that use difference GMM, the independent variables of interest are assumed to be endogenous. In fact, the main purpose of my difference GMM robustness checks is to control for the potential (and likely) endogenous relationship between payout policy and institutional ownership.

My implementation of difference GMM starts with the following basic model which will be transformed by the difference GMM process.

*Policyit*

*Policyit* 1 *Instit*1 *Controlit*1 *i*  *it*

(1)

In this model, *Policy*it represents the change in the firm payout (or repurchases or payout composition) policy. *Policy*it-1 represents the change in firm payout policy in the previous year. The independent variable *Inst*i*t-1* represents the change in institutional ownership percentage in the previous year. Controli*t-1* represents a vector of time-varying firm level control variables. Year dummies are included as control variables to remove time-related shocks that affect all firms. Firm-specific (fixed effects) errors and time-varying observation-specific errors are represented by νi and εit, respectively.

Several econometric problems which are endemic to model C-1 can be corrected by difference GMM. The change in institutional ownership percentage (*Inst*i*t-1*) is assumed to be endogenous. Therefore, it is instrumented with lagged changes in institutional ownership. This predetermines the institutional ownership variable thus rendering it uncorrelated with the error term. Similarly, the use of the lagged dependent variable (*Policy*it-1) as an independent variable leads to autocorrelation. This variable is also instrumented with lags of itself. Firm fixed effects are contained in the error term νi. The difference GMM methodology uses first-differences to transform model C-1 thus removing the firm fixed effects error term because it is time invariant. The new model is shown in model 2.